

06 de outubro de 2014



Universidade do Minho
Serviços de Documentação

Estudo sobre os Dados Científicos gerados no âmbito da investigação produzida na Universidade do Minho

**Pedro Príncipe, Ricardo Saraiva, Clara Boavida, Raquel Truta,
Eloy Rodrigues, José Carvalho | openaccess@sdum.uminho.pt**

Tópicos da apresentação

1. Motivação e enquadramento: gestão de dados científicos na Universidade do Minho
2. Âmbito e caracterização do estudo desenvolvido
3. Resultados do estudo sobre curadoria e gestão de dados científicos
4. Resultados do estudo sobre Acervo de dados científicos existente
5. Conclusões e trabalho futuro



1

MOTIVAÇÃO & ENQUADRAMENTO



Motivação & enquadramento

- A gestão de dados científicos tem assumindo **crescente relevância no contexto das instituições**.
 - **reforçar** a sua capacidade de gestão e curadoria dos seus recursos para **apoiarem** as **necessidades de dados** dos investigadores.
- **Requisitos dos financiadores de ciência:**
 - **Comissão Europeia (CE)** criou recentemente, no âmbito do programa quadro Horizonte 2020, um **projeto-piloto de dados abertos**.
 - A **FCT** em Portugal adotou no contexto da sua política de Acesso Aberto, **orientações sobre a disponibilização de dados de projetos com seu financiamento**.





OBJ.3

Desenvolver um programa para a criação de serviços de gestão de dados científicos na UMinho

Ind. 3.1 Diagnóstico dos dados científicos na UMinho

Ind. 3.2

Conteúdos/Serviços de apoio e consultoria de gestão dados científicos

Ind. 3.3 Gestão dos dados produzidos/recolhidos nos SDUM

Plano de intervenção dos SDUM

- Resulta de um trabalho de análise da equipa de projetos Open Access
- Define **10 eixos de intervenção**
- Enquadrados em **4 áreas estratégicas**
- **33 ações programadas para 2014 e 2015**



Áreas estratégicas do plano de intervenção:



Universidade do Minho
Serviços de Documentação

Diagnóstico & política

Serviços de suporte & informação

Infraestruturas & normalização

Competências & recursos humanos

Poster ConfOA 2014





Eixos de intervenção

- 1. Análise da realidade da gestão dos dados científicos produzidos na UMinho.**
- 2. Serviços de apoio e consultoria na gestão de dados científicos.**
- 3. Especificação de um serviço interno dedicado para dados científicos e desenvolvimento de competências dos colaboradores SDUM.**
- 4. Desenvolvimento da política institucional para a gestão de dados científicos.**
- 5. Promoção da interoperabilidade de infraestruturas de dados científicos.**
- 6. Serviços de curadoria para o ciclo de vida dos dados científicos.**
- 7. Serviços de metadados para dados de investigação.**
- 8. Promoção da referenciação e citação dos dados científicos.**
- 9. Serviço de repositório de dados científicos.**
- 10. Melhorar a prática da gestão dos dados científicos e a disponibilização de dados aberto.**

Inquérito sobre Dados Científicos

Inquérito sobre Dados Científicos

Este inquérito por questionário pretende recolher informações com vista a **dados científicos** (*research data / scientific data*) que são gerados no âmbito da

Para o propósito deste estudo, considera-se *research data* todos os dados utilizados no decorrer da sua investigação na Universidade do Minho. Os dados podem ser primários recolhidos por si ou pelo seu grupo de investigação ou dados secundários fornecidos por

O questionário demorará cerca de 10 minutos a ser respondido.

Agradecemos desde já a sua colaboração! Para esclarecimentos adicionais

Existem 53 pessoas



Este inquérito por questionário pretende recolher informações com vista a proporcionar um melhor conhecimento sobre os dados científicos ('research data' / 'scientific data') que são gerados no âmbito da investigação produzida na Universidade do Minho. Para o propósito deste estudo, considera-se *research data* todos os dados que atualmente possui, que foram recolhidos e/ou utilizados no decorrer da sua investigação na Universidade do Minho. Os dados científicos podem ser dados primários recolhidos por si ou pelo seu grupo de investigação ou dados secundários fornecidos por terceiros. O questionário demorará cerca de 10 minutos a ser respondido. Agradecemos desde já a sua colaboração! Para esclarecimentos adicionais, por favor, contacte: openaccess@sdum.uminho.pt

Section A: A. Identificação

A1. Unidade Orgânica de Ensino e de Investigação (UOEI)

- Escola de Arquitectura (EA) ☐
- Escola de Ciências (EC) ☐
- Escola de Ciências da Saúde (ECS) ☐
- Escola de Direito (ED) ☐
- Escola de Economia e Gestão (EEG) ☐
- Escola de Engenharia (EE) ☐
- Escola de Psicologia (EP) ☐
- Escola Superior de Enfermagem (ESE) ☐
- Instituto de Ciências Sociais (ICS) ☐
- Instituto de Educação (IE) ☐
- Instituto de Letras e Ciências Humanas (ILCH) ☐
- Outro ☐

A2. Categoria Profissional

- Docente Universitário ☐
- Investigador ☐
- Técnico Superior ☐
- Bolsista(a) de Investigação ☐



2

ÂMBITO & CARACTERIZAÇÃO



Âmbito e caracterização



Universidade do Minho
Serviços de Documentação

- Inquérito por questionário:
 - **Anónimo** e composto por **53 perguntas**
 - Disponibilizado **via web** usando o software **LimeSurvey**
 - Dirigido a investigadores, docentes e centros de investigação da UMinho
- Divulgado exclusivamente através de **e-mail enviado pelo Vice-Reitor** responsável pela área da investigação.
- Período de resposta : **14 de março e 17 de abril de 2014.**

226 respostas válidas
de um universo total de 353 respostas.

Estrutura do inquérito



Universidade do Minho
Serviços de Documentação

Este inquérito por questionário pretende recolher informações com vista a proporcionar um melhor conhecimento sobre os dados científicos ('research data' / 'scientific data') que são gerados no âmbito da investigação produzida na Universidade do Minho. Para o propósito deste estudo, considera-se *research data* todos os dados que atualmente possui, que foram recolhidos e/ou utilizados no decorrer da sua investigação na Universidade do Minho. Os dados científicos podem ser dados primários recolhidos por si ou pelo seu grupo de investigação ou dados secundários fornecidos por terceiros. O questionário demorará cerca de 10 minutos a ser respondido. Agradecemos desde já a sua colaboração! Para esclarecimentos adicionais, por favor, contacte: openaccess@s dum.uminho.pt

Section A: A. Identificação

A1. Unidade Orgânica de Ensino e de Investigação (UOEI)

- Escola de Arquitectura (EA) ☐
- Escola de Ciências (EC) ☐
- Escola de Ciências da Saúde (ECS) ☐
- Escola de Direito (ED) ☐
- Escola de Economia e Gestão (EEG) ☐
- Escola de Engenharia (EE) ☐
- Escola de Psicologia (EP) ☐
- Escola Superior de Enfermagem (ESE) ☐
- Instituto de Ciências Sociais (ICS) ☐
- Instituto de Educação (IE) ☐
- Instituto de Letras e Ciências Humanas (ILCH) ☐
- Outro ☐

Outro

--	--	--	--	--	--	--	--	--	--

A2. Categoria Profissional

- Docente Universitário ☐
- Investigador ☐
- Técnico Superior ☐
- Bolsista(a) de Investigação ☐

IDENTIFICAÇÃO
(CARACTERIZAÇÃO DOS PARTICIPANTES)

CURADORIA E GESTÃO DE DADOS CIENTÍFICOS

ACERVO DE DADOS CIENTÍFICOS EXISTENTES

Caracterização dos participantes

- **Respostas de 10 das 11 unidades** existentes na UMinho;
- **Metade dos participantes pertence à Escola de Engenharia** (55,8%) seguindo-se a Escola de Ciências com 13,7%;
- Categoria profissional, identificaram-se: **54,9% docentes universitários, 27,0% bolseiros e 20,4% investigadores;**
- Perfil da atividade: **52,2% são participantes em projetos, 35% coordenadores de projetos, 15,5% alunos de doutoramento;**
- Representatividade da **amostra**, regista-se a participação de um universo aproximadamente **correspondente a 20% do pessoal docente e investigador da UMinho.**



3

CURADORIA E GESTÃO DE DADOS CIENTÍFICOS





3. Curadoria e gestão de dados científicos (1/9)

	<i>N</i>	%
Sim	210	92,9%
Não	16	7,1%
	226	

No seu trabalho como investigador ou dos projetos em que participa **produz ou possui dados científicos?**

93% SIM



3. Curadoria e gestão de dados científicos (2/9)

	<i>N</i>	%
Experimentais	157	74,8%
Estatísticos	100	47,6%
Imagens	87	41,4%
Numéricos	76	36,2%
Observacionais	71	33,8%
Computacionais	64	30,5%
Textuais	63	30,0%
Multimédia	17	8,1%
Áudio	15	7,1%
Outro	5	2,4%
	210	

Que tipos de dados científicos são produzidos na investigação que realiza / em que participa?

75% Experimentais . 48% Estatísticos . 41% Imagens



3. Curadoria e gestão de dados científicos (3/9)

	<i>N</i>	%
Menor que 1 GB	18	8,6%
1 - 10 GB	51	24,3%
10 - 100 GB	44	21,0%
100 - 1 TB	25	11,9%
1 - 10 TB's	11	5,2%
10 - 100 TB's	4	1,9%
Mais de 100 TB's	5	2,4%
Não sei	52	24,8%
	210	

Qual é o **volume global aproximado de dados acumulados** por si ou pelo seu grupo de investigação?

25% não consegue estimar

24% entre 1-10GB

21% entre 10-100GB



3. Curadoria e gestão de dados científicos (4/9)

	<i>N</i>	%
Disco externo	136	80,0%
Disco de PC	89	52,4%
Dropbox	55	32,4%
Servidor local	40	23,5%
USB / Flash drive	37	21,8%
Disco de equipamento	29	17,1%
CD / DVD	14	8,2%
Outro	8	4,7%
Servidor internacional	3	1,8%
Servidor nacional	2	1,2%
Tape	2	1,2%
	170	

Onde são guardados os dados?

80% disco externo

52% disco de PC

32% dropbox

3. Curadoria e gestão de dados científicos (5/9)

	<i>N</i>	%
Não	169	80,5%
Sim	41	19,5%
	210	

Possui algum plano para a gestão de dados resultantes da sua investigação (por exemplo, política de preservação de dados, política de gestão de registos, estratégia de eliminação de dados)?

80% NÃO



3. Curadoria e gestão de dados científicos (6/9)

	<i>N</i>	%
Requisito da investigação para aceder, analisar e consultar dados de outros	18	43,9%
Tamanho da equipa do projeto (ou seja, vários criadores e utilizadores de dados)	6	14,6%
Volume de dados associados ao projeto	6	14,6%
Exigência do financiador do projeto	3	7,3%
Complexidade de dados associados com o projeto (por exemplo, vários formatos)	3	7,3%
Outro	3	7,3%
Ausência de política de gestão de dados no grupo de investigação	2	4,9%
	41	

Qual foi o **principal motivo** para o **desenvolvimento de um plano de gestão de dados?**

44% para aceder, analisar e consultar outros dados

15% tamanho da equipa

15% volume de dados associados ao projeto



3. Curadoria e gestão de dados científicos (7/9)

	<i>N</i>	%
Acesso restrito ao grupo / projeto de investigação	130	57,5%
Acesso restrito ao investigador	35	15,5%
São todos de acesso aberto imediato	30	13,3%
São de acesso aberto após período de embargo (por favor especifique a duração na caixa de texto)	23	10,2%
Outro	8	3,5%
	226	

No que concerne ao tipo de acesso, qual é o estatuto usual dos seus dados?

57% acesso restrito ao grupo ou projeto

25% acesso aberto imediato ou após embargo



3. Curadoria e gestão de dados científicos (8/9)

	<i>N</i>	%
Confidencialidade / direitos de propriedade intelectual	166	79,0%
Potencial reutilização	93	44,3%
Possível má interpretação dos dados	65	31,0%
Tempo / esforço despendido na sua obtenção	46	21,9%
Valor comercial dos dados	41	19,5%
Outro	10	4,8%
	210	

**Que tipo de questões o preocupa
relativamente ao seu acesso aberto?**

79% Confidencialidade / propriedade intelectual

44% potencial reutilização



3. Curadoria e gestão de dados científicos (9/9)

	<i>N</i>	%
Não	176	83,8%
Sim	34	16,2%
	210	

Tem conhecimento do projeto-piloto de dados abertos da Comissão Europeia no âmbito do programa quadro H2020?

84% NÃO

16% SIM



4

ACERVO DE DADOS CIENTÍFICOS EXISTENTE





4. Acervo de dados científicos existentes (1/4)

	<i>N</i>	%
2 - 5 anos	56	26,7%
5 - 10 anos	56	26,7%
Indefinido	47	22,4%
10 - 20 anos	18	8,6%
Não sei	16	7,6%
1 - 2 anos	10	4,8%
20 - 100 anos	5	2,4%
Menos de 1 ano	1	0,5%
Sem resposta	1	0,5%
Outro	0	0,0%
	210	

Período de preservação estimado

27% 2-5 anos

27% 5-10 anos

22% indefinido



4. Acervo de dados científicos existentes (2/4)

	<i>N</i>	%
1 - 10 GB	72	34,3%
10 - 100 GB	63	30,0%
100 GB - 1 TB	30	14,3%
Menos de 1 GB	23	11,0%
1 - 10 TB's	10	4,8%
Sem resposta	8	3,8%
Múltiplos TB's	4	1,9%
	210	

**Dimensão 'final' estimada dos dados
(tendo por exemplo um projeto tipo)**

34% 1-10 GB

30% 10-100 GB

14% 100 GB - 1 TB



4. Acervo de dados científicos existentes (3/4)

	<i>N</i>	%
Dados brutos obtidos através de instrumentos	145	69,0%
Documento de texto (ex. Word, PDF, ...)	145	69,0%
Folha de cálculo (ex. Excel, ...)	142	67,6%
Imagens, digitalizações ou raios x	124	59,0%
Dados brutos gerados por programas	112	53,3%
Notas de laboratório	97	46,2%
Ficheiros SPSS	81	38,6%
Bases de dados (ex. MySQL, Oracle, ...)	80	38,1%
Vídeo	61	29,0%
Software	55	26,2%
Dados clínicos	53	25,2%
Áudio	42	20,0%
Outro formato proprietário	25	11,9%
Outro	5	2,4%
	210	

Que tipo de dados deverão ser preservados?

69% dados brutos de instrumentos

69% documentos de texto

68% folhas de cálculo



4. Acervo de dados científicos existentes (4/4)

	<i>N</i>	%
Sistema de backup de ficheiros	153	72,9%
Serviços de armazenamento e preservação	136	64,8%
Apoio técnico na elaboração de planos de gestão de dados	84	40,0%
Formação e consultoria na gestão de dados científicos	69	32,9%
Serviços de partilha e acesso a dados	42	20,0%
Informação sobre repositórios temáticos de dados	33	15,7%
Identificação de normas e esquemas disciplinares de descrição de dados	14	6,7%
Outro	6	2,9%
	210	

Que tipo de serviços gostaria que a UMinho fornecesse para garantir o acesso a esses dados em termos futuros?

73% sistema de backup de ficheiros

65% serviços de armazenamento e preservação

40% apoio técnico na elaboração de Planos de GD



5

CONCLUSÕES & TRABALHO FUTURO



Conclusões e trabalho futuro

- Permitiu obter dados para **consolidar e ajustar o plano de intervenção** para serviços de suporte à Gestão dos Dados Científicos na UMinho.
- As mais de 200 respostas recolhidas permitem obter uma **melhor panorâmica do tipo de dados produzidos** no âmbito da investigação realizada na UMinho.
- A **grande diferenciação** entre as Unidades Orgânicas de Ensino e Investigação no nível de participação do estudo é **um indicador de diferentes graus de importância e de sensibilização**.
- Os resultados indiciam **abertura para a disponibilização em breve de serviços, mas implicam aprofundar com comunidades piloto** o seu desenvolvimento.

Conclusões e trabalho futuro

- **Existe necessidade, espaço e abertura na comunidade da UMinho para o desenvolvimento, implementação e disponibilização de serviços e ferramentas associadas à gestão de dados científicos, em particular:**
 - ✓ Serviços de armazenamento e backup de dados;
 - ✓ Serviços de apoio e consultoria, com particular enfoque nas ferramentas de suporte ao desenvolvimento de planos de gestão de dados;
 - ✓ Serviços de repositório de dados científicos, possibilitando o armazenamento, referência, acesso e reutilização de dados a longo prazo.

Conclusões e trabalho futuro

- Os serviços já em realização pelos SDUM (suporte da **ferramenta DMPonline para planos de gestão de dados; guias (in)formativos e sessões de formação interna...**) e os projetos a desenvolver proximamente na Universidade do Minho serão:
 - ✓ em consonância com as orientações da Reitoria,
 - ✓ em articulação com outros Serviços da UMinho,
 - ✓ em parceria com parceiros relevantes fora da UMinho.



Relatório e dados disponíveis no RepositóriUM:

Rodrigues, E., Boavida, C., Truta, R., Príncipe, P., Carvalho J. & Saraiva, R. (2014). Análise dos resultados do inquérito sobre dados científicos produzidos na Universidade do Minho. Universidade do Minho, Braga. Disponível no RepositóriUM em: <http://hdl.handle.net/1822/29948>

RELATÓRIO DO INQUÉRITO

<http://hdl.handle.net/1822/29948>



DADOS DO INQUÉRITO

<http://hdl.handle.net/1822/29949>



Obrigado pela atenção!



Universidade do Minho
Serviços de Documentação

[HTTP://OPENACCESS.SDUM.UMINHO.PT](http://openaccess.sdum.uminho.pt)

Clara Boavida, Eloy Rodrigues, José Carvalho,
Pedro Príncipe, Raquel Truta e Ricardo Saraiva

openaccess@sdum.uminho.pt

